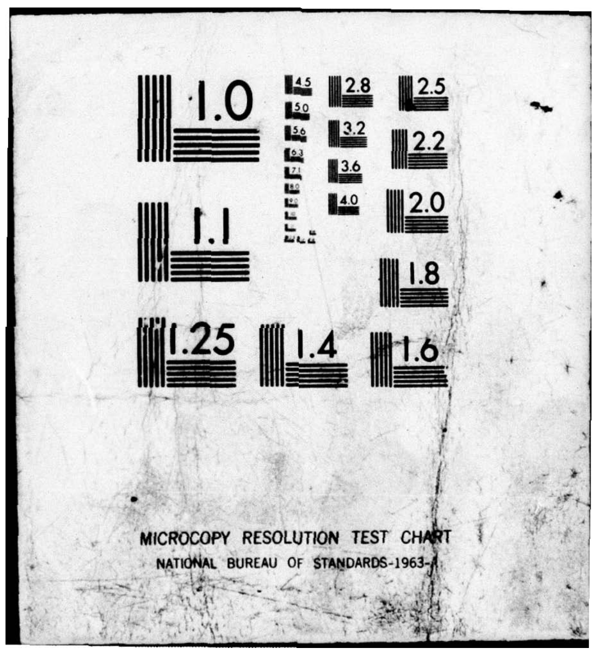
AD-A050 910 MILITARY ACADEMY WEST POINT N Y DEPT OF ENGINEERING F/G 6/5
A DECISION ANALYTIC APPROACH TO CORONARY CARE PATIENT CLASSIFIC--ETC(U)
OCT 76 T A BRESNICK

UNCLASSIFIED NL

1 OF 1
AD
A050 910

END
DATE
FILMED
4-78
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A DECISION ANALYTIC APPROACH TO CORONARY CARE PATIENT CLASSIFICATION

CAPTAIN TERRY A. BRESNICK

DEPARTMENT OF ENGINEERING
UNITED STATES MILITARY ACADEMY
WEST POINT, NEW YORK 10996

OCTOBER 1976

ABSTRACT

The purpose of this project is to categorize cardiac patients using an expected utility criteria. Bayesian updating using symptomatic information has been a successful tool for disease diagnosis. However, in classification of cardiac patients, discriminant analysis using the BMDO5M computer package achieves comparable results with far less effort. Discriminant scores can be evaluated and normalized as probabilistic estimates of the patient's cardiac condition. By incorporating these probabilities with subjective value estimates for possible misclassifications, the expected utility of assigning patients to each cardiac category can be calculated. Patients are then classified into cardiac categories based upon highest expected utility.

A

B

ACKNOWLEDGMENTS

I would like to express my thanks to several people: Dr. John S. Schroeder of the Cardiology Division at the Stanford University Medical Center for his medical expertise; Dr. Edward J. Sondik of the Engineering Economic Systems Department of Stanford for his outstanding guidance and assistance; Colonel Allen F. Grum, Department of Engineering, United States Military Academy for supporting my efforts; and Miss Joyce Wall and Mrs. Donna Lord for their assistance in preparing this report.

I. INTRODUCTION

Typically, patients with chest pains appear at coronary units and are faced with a barrage of medical histories, testing procedures, and lengthy hospital stays before their specific problem can be isolated. Often, patients suffer further deterioration in health before treatment actually begins. There are varied symptoms to examine, many of which provide minimal information. There is a wide selection of tests and experiments that can be performed, often at a high cost to the patient in terms of both money and time. Many of the tests provide little additional diagnostic information.

This paper provides physicians at coronary care units with a non-invasive model for classification of cardiac patients. By applying the techniques of discriminant analysis and expected utility decision making, patients can be diagnosed upon admission to the coronary care unit (CCU), using no more than medical history, prodromal data, and routine testing procedures. One approach to this problem involves calculating the probabilities that a patient has a specific cardiac condition, evaluating the consequences of treating a misclassified patient, and making a diagnosis that would maximize the "benefits" to the patient. The basic approach thus identifies patient categories for treatment more rapidly and at less expense to the patient. This in turn leads to higher quality medical care.

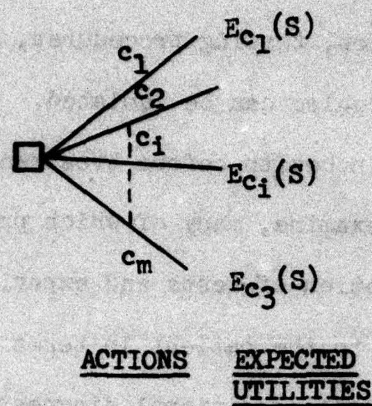
In general mathematical terms, we can describe the decision problem for a specific patient as

$$\max_{c} \quad E_c(S) \tag{1}$$

where

$$c \in [c_1, c_2, \ldots, c_m]$$

and $E_c(S)$ is the expected utility of placing a patient with symptoms S into a category c. This can be schematically represented as



$$E_{c_i}(S) = \sum_{j=1}^{m} \text{Prob}\,\{\hat{c}_j \mid S\}\, u_{ji}$$

Figure 1

where

c_i is the diagonosed category

$\hat{c}_j$ is the "true" category

u_{ji} is the utility of classifying and treating a true category j as a category i.

This general approach will be described in detail in the paper.

II. BACKGROUND

In 1974, the Cardiology Division at Stanford University Medical Center published a procedure for analyzing the clinical course of chest pain admissions. The purpose was to identify coronary disease patients prior to the actual incidence of infarction. Examination of medical histories and prodromal data could lead to detection of potential infarctions, and action could be taken to avoid the sudden death that frequently accompanies infarction.[1]

John S. Schroeder, M.D., Alfred P. Spivak, M.D., and Irene Lamb, R.N., all of the Cardiology Division, included the following goals in their analysis:

A. To determine the incidence and frequency of patients admitted to the coronary care unit or coronary surveillance unit who have had:

1. Myocardial infarction (M.I.) ruled out, chest pain not of cardiac origin.

2. Myocardial infarction ruled out, unstable angina present.

3. Subendocardial infarction.

4. Definite myocardial infarction.

5. Pre-infarction angina with progression to myocardial infarction during hospitalization.

B. To characterize and compare prodromal symptoms of these patient groups.

C. To identify environmental or psycho-social stresses which may have contributed to the onset of chest pain.

D. To identify clinical and laboratory characteristics of the pre-infarction patient who will progress to myocardial infarction or death during hospital stay.

E. To identify patients with high risk of subsequent myocardial infarction or death following hospital discharge.

This project is primarily concerned with building a model to accomplish their first goal. The coronary categories that they established will be used here.

III. METHODOLOGY

In establishing a model, our efforts followed the steps given below:

1. Data collection.
2. Assessment of a utility structure.
3. Selection of the key variables.
4. Construction of classification model based on key variables.
5. Validation of the model.

DATA COLLECTION

Dr. Schroeder provided the data on 125 coronary care unit patients. The data included historical information, prodromal symptoms, admission symptoms, laboratory results, etc., as well as a retrospective determination of the disease category. We considered this final determination to be the "true" patient state. After eliminating two patients whose data appeared to contain serious outliers, the following breakdown of patients remained:

CATEGORY	NUMBER OF PATIENTS
1. M.I. ruled out, non-cardiac pain	15
2. M.I. ruled out, unstable angina	55
3. Subendocardial infarction	10
4. Definite myocardial infarction	29
5. Pre-infarction, progress to M.I.	14

To insure that additional data would be available to later test the diagnostic model, twenty three patients records were randomly removed from the data base. This left 100 patients available for analysis.

ASSESSMENT OF A UTILITY STRUCTURE

Methods of classifying patients into disease categories using probability assessments are well documented. However, high probabilities of a

disease category are often insufficient to establish an optimal diagnosis. Clearly, it is far more serious to treat an infarction patient as a non-cardiac pain victim, than it is to treat a non-cardiac pain patient as an infarction victim. We therefore looked at the value or utility of classifying patients into the various categories. Since Dr. Schroeder was the decision maker for this project, we used his utility structure.

This study used both a direct approach and a lottery approach to evaluate the relative utility of the possible classification schemes. In describing the classifications, we used the coronary categories specified earlier. In the direct approach, Dr. Schroeder subjectively rank ordered the types of classifications from best to worst. The best outcome would be to correctly identify a category 1 patient, while the worst would be to place a category 4 patient into category 1. By arbitrarily placing a value of +100 on the best case, and a value of -100 on the worst case, Dr. Schroeder evaluated all of the possibilities on a relative basis.

In the lottery approach, we used expected utility theory.[2] This implies that if an individual is satisfied with either receiving a guaranteed award, or taking part in a lottery, then the utility of the sure award equals the expected utility of the lottery. Dr. Schroeder faced a series of lotteries similar to those shown in Figure 2. He was presented with a guaranteed classification, such as placing a true category 1 into category 5, and two probabilistic outcomes, such as placing a true category 1 into category 1 and a true category 4 into category 1. He had to assess the probabilities that would make him indifferent to the alternatives. In other lotteries, he was given the possible outcomes and the associated probabilities, and he had to select a sure outcome that would achieve

LOTTERY

guaranteed outcome		possible outcomes and probabilities	values
		Put Cat. 1 → Cat. 1 (.8)	+100
Put Cat. 1 → Cat. 5	~		
		Put Cat. 4 → Cat. 1 (.2)	-100

CALCULATIONS

$$\begin{aligned} U\,(\text{Cat. } 1 \rightarrow \text{Cat. } 5) &= .8\,U\,(\text{Cat. } 1 \rightarrow \text{Cat. } 1) + .2\,U\,(\text{Cat. } 4 \rightarrow \text{Cat. } 1) \\ &= .8\,(100) + .2\,(-100) \\ &= 60 \end{aligned}$$

LOTTERY

guaranteed outcome		possible outcomes and probabilities	values
		Put Cat. 1 → Cat. 1 (.9)	+100
Put Cat. 1 → Cat. 2	~		
		Put Cat. 1 → Cat. 5 (.1)	60

CALCULATIONS

$$\begin{aligned} U\,(\text{Cat. } 1 \rightarrow \text{Cat. } 2) &= .9\,U\,(\text{Cat. } 1 \rightarrow \text{Cat. } 1) + .1\,U\,(\text{Cat. } 1 \rightarrow \text{Cat. } 5) \\ &= .9\,(100) + .1\,(60) \\ &= 96 \end{aligned}$$

Figure 2

indifference. We compared the results of the lottery evaluation method with the direct method, and we analyzed areas of discrepancy. By varying outcomes and probabilities, we insured that Dr. Schroeder's choices were consistent. The final utility, or benefit, matrix, $[U]$, is given in Appendix 1. The components of this matrix, u_{ji}, represent the utility of treating a true category j patient as a category i patient.

It is important to stress that the values in the utility matrix are Dr. Schroeder's alone. This might be appropriate since he is the decision maker in this case. Each decision maker will have his own utility structure which will be a function of his experience, environment, and attitudes. This point has been one of great controversy. Is it the doctor's or the patient's utility structure that is most appropriate? Perhaps the doctor should use his experience to help place the patient's utilities in perspective? We made no attempt to argue this question in our paper. We used Dr. Schroeder's utility matrix which could then be subjected to sensitivity analysis. With this understanding, we can later use the utility matrix in conjunction with the probabilities that a patient actually falls within a category to calculate the expected utility of classification.

SELECTION OF THE KEY VARIABLES

Since we collected data on more than 100 signs or symptoms, it was clear that some pruning was necessary.[3] We concluded that many of the data items were insignificant and we eliminated them from initial consideration. We selected less than 30 "key" symptoms, and this was served as an appropriate starting point. These symptoms are included at Appendix 2.

As mentioned previously, both probabilistic assessment and value assessment are required for an expected utility model. By applying the B.M.D. Biomedical Data Programs developed at the Health Sciences Computing Facility at U.C.L.A., we could estimate posterior probabilities for each category. In particular. BMD05M (Discriminant Analysis for Several Groups) was used. This program computes linear discriminant functions for classification into one of several groups.[4]

We can select an m-component vector of symptoms, $\underline{Y}$, for classification, and we assume that the underlying distribution is multivariate normal. We want to classify a random patient k with symptom vector $\underline{Y}$ into one of the five coronary categories 1, 2, ... 5. We further assume that the symptoms for each category are distributed $N(\underline{\mu}_i, \underline{\Sigma}_i)$, and that $\underline{\Sigma}_1 = \underline{\Sigma}_2 = \ldots = \underline{\Sigma}$.[5]

Using the sample parameters $\underline{X}_i$ and $\underline{S}_i$ as estimates of the population parameters $\underline{\mu}_i$ and $\underline{\Sigma}_i$, BMD05M performs the following computational steps:

Required transgenerations on the data

Means of symptoms for each category

Sum of products of deviation from the mean

Pooled dispersion matrix and its inverse

Mahalanobis distance

Coefficients α_{im} and constant term α_{i0} for each category i

Using these α's, the discriminating function for the i^{th} category is calculated as:

$$f_i(\underline{Y}_k) = \sum_{j=1}^{m} y_{kj}\alpha_{ij} + \alpha_{i0} \tag{2}$$

where m is the number of symptoms, i is the category number, and k is the patient number.

The discriminant scores are normalized to represent probabilities for each category as follows:

$$\text{Prob}\,\{i \mid \underline{Y}_k\} = \frac{e^{(f_i(\underline{Y}_k) - f(\underline{Y}_k)\,\text{Max})}}{\sum_{n=1}^{5} e^{(f_n(\underline{Y}_k) - f(\underline{Y}_k)\,\text{Max})}} \qquad (3)$$

The typical output of BMD05M includes these probabilities as well as a table of classifications based upon largest probability.

To classify patients based upon a utility criteria rather than upon a largest probability, we modified BMD05M to calculate the expected utilities of classifying each patient k into each category i. Using the utility matrix, [U], and the probabilities from the discriminant analysis, we calculated the expected utility of classifying patient k into category i as:

$$E_i(\underline{Y}_k) = \sum_{n=1}^{5} \text{Prob}\,\{n \mid \underline{Y}_k\}\, u_{ni} \qquad (4)$$

Classification is then based upon the highest expected utility. We then calculated the average expected utility for all categories i, (the entire 100 patient data base) as:

$$E(\text{average}) = \Big(\sum_{k=1}^{100} \sum_{n=1}^{5} \text{Prob}\{n \mid \underline{Y}_k\}\, u_{ni}\Big)/100 = \sum_{k=1}^{100} E_i(\underline{Y}_k)/100 \qquad (5)$$

The procedure is schematically portrayed in Figure 3.

THE CLASSIFICATION MODEL

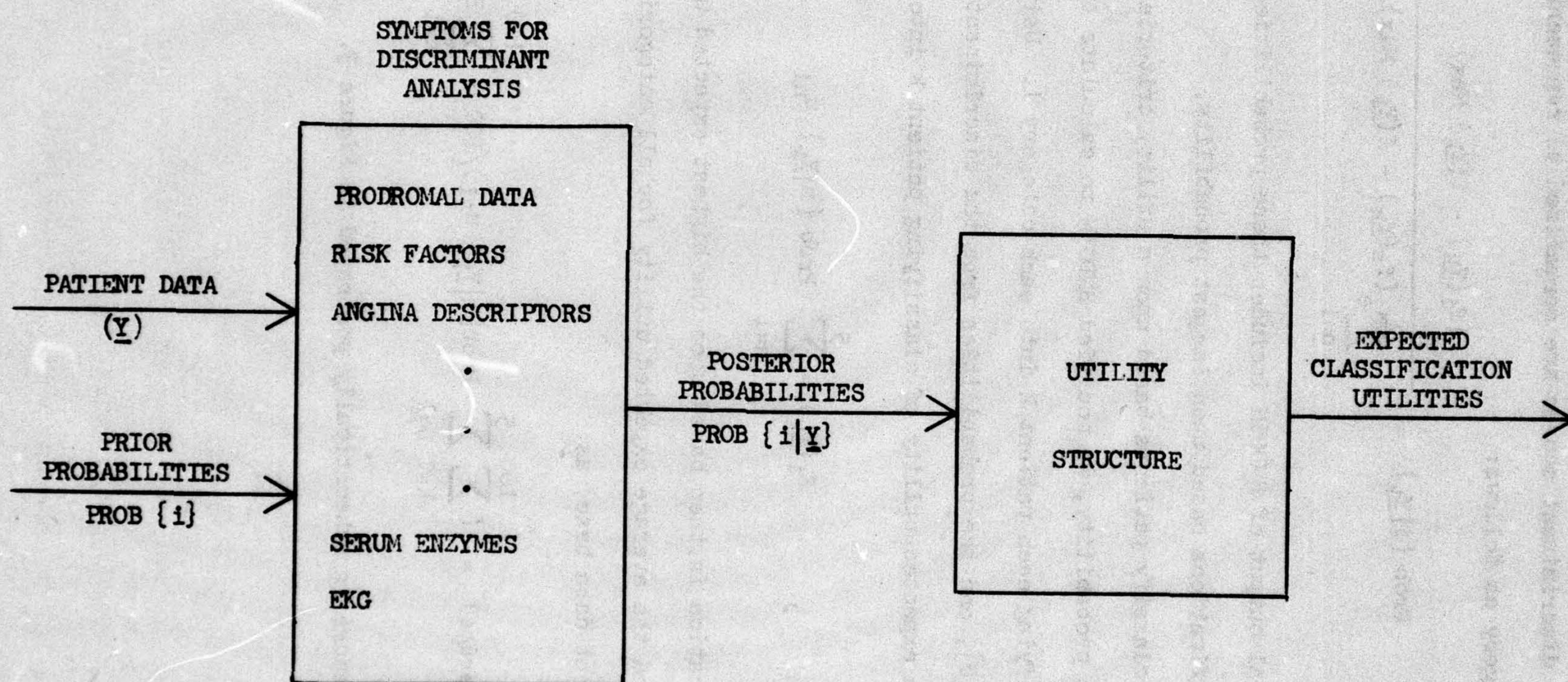


Figure 3

The problem now remained to determine which symptoms provided the most useful discrimination. We examined many combinations of symptoms and calculated the average expected utility using each symptom set. We found that risk factors and prodromal data generally were comparable for all categories and added little to the ability to discriminate. The most discriminating symptoms were found in the following information:

Unstable angina descriptions

Admission chest pain descriptions

Electrocardiograph (EKG)

ST. segment and shift

T-wave and old Q-wave

Enzyme shift for serum glutamic oxaloacetic transaminase (SGOT)

Enzyme shift for creatine phosphokinase (CPK)

While the other symptoms certainly provided some information, they provided no marginal assistance in our ability to distinguish between categories.

These observations established a 14 component symptom vector. We could now use this vector as a basis for classifying new patients. It is important to note that classifying category 4 patients is a simple task. EKG, unstable angina, and serum enzymes are sufficient for an extremely reliable diagnosis. Category 4 patients are rarely misdiagnosed. As a result, we were more interested in a scheme of classifying the other categories of patients.

CONSTRUCTION OF THE MODEL BASED ON KEY VARIABLES

Using the vector of key symptoms and the data base information, we obtained the most useful set of α's for each coronary category. We then developed a computer algorithm to evaluate new patients. The vectors provided the discriminant scores for each patient q for each category i.

This was:

$$f_i(\underline{Y}_q) = \sum_{j=1}^{14} y_{qj}\alpha_{ij} + \alpha_{i0} \qquad (6)$$

We used this in turn to calculate the probabilities and expected utility for each category as follows:

$$E_i(\underline{Y}_q) = \sum_{n=1}^{5} \frac{e^{(f_i(\underline{Y}_q) - f(\underline{Y}_q)\ \max)}}{\sum_{\ell=1}^{5} e^{(f_\ell(\underline{Y}_q) - f(\underline{Y}_q)\ \max)}} \qquad (7)$$

The model was now complete for diagnosing any future cardiac patients.

<u>VALIDATION OF THE MODEL</u>

Using the discriminant coefficients (α) obtained for the most discriminating symptoms, we now had to determine if our approach would work on additional patients. By applying the model to the 23 patients removed from the original data base, we calculated the discriminant scores, probabilities, and expected classification utilities of classifying each patient into each category. Diagnosis was based upon highest expected utility. Twenty-three patients are far too few observations to make firm conclusions, but they allowed us to draw informal inferences about the validity of the model.

If all patients had been correctly categorized, we would have achieved an expected average classification utility of 85. We actually achieved value of 77.6. We found this to be most encouraging. Of the 23 patients, we found two instances of what can be called serious misclassifications. These represent patients that were classified into categories less severe in nature than their true category. We compared

these results with classification based upon largest probability, and found five serious misclassifications using the latter approach. We also found several misdiagnoses in which patients were placed into a category more severe than their true category, but these are far more tolerable. They represent instances in which the patient receives extra care.

Although 23 patients provided insufficient information to place complete confidence in the model, the results we achieved were far too successful to invalidate it. It is clear that we must now obtain additional data before we can make suitable statistical inferences about our results.

There are other modifications to the model that can also be examined in the future. We can perform sensitivity on the utility matrix and explore the question of whose preference structure should be used. This model dealt with non-invasive symptom input, but we might look at other procedures such as angiography. Although an angiogram is considered to be a surgical procedure, the benefits gained might exceed the difficulties associated with it. Additionally, we could extend the general model to other areas of medical practice.

CONCLUSIONS

The expected utility classification model, using discriminant analysis and subjective value assessment, is a worthwhile tool for coronary care patient classification. It provides the decision maker with a rapid, inexpensive method of classifying patients without invasive procedures. The model is a recursive one that is continuously updated and refined as new patient data becomes available. The approach is flexible and is designed to fit the decision maker's expertise and the appropriate preference structure.

Initial efforts in applying the model to actual diagnosis have proven successful and provide the impetus to expand the data base. This should make the model more reliable and allow doctors to use it with confidence. This in turn, will provide the patient with higher quality medical care.

APPENDIX 1

Table of Utilities for Placing Patients with

True Category i into Diagnosed Category i*

		DIAGNOSED CATEGORY i*				
		1	2	3	4	5
TRUE CATEGORY i	1	100	96	78	52.5	60
	2	49.8	98	65.8	51.2	60.4
	3	-68.6	-33	86	64	67.4
	4	-100	-62.8	47	58.1	50
	5	-90	-42.4	60	71.8	73.4

APPENDIX 2

Table of Symptoms (After Pruning)

Age

Sex

Number of months known heart disease

Risk factors

- hypertension
- obesity
- hyperlipids
- diabetes mellitus
- smoking
- family history of heart disease

Prodromal symptoms

- generalized
- cardiovascular
- gastrointestinal
- personality

Unstable angina description

Admission chest pain description

EKG

ST. segment and shift

T wave and Q wave

Serum enzyme shifts

- SGOT
- CPK

ENDNOTES

[1]Dr. John S. Schroeder, Irene Lamb, and Dr. Alfred P. Spivack, "Clinical Course of Chest Pain Admissions," (Cardiology Division of Stanford University Medical Center unpublished report, October 1974), pp. 1-6.

[2]Allen S. Ginsberg and Fred Offensend, "An Application of Decision Theory to a Medical Diagnosis - Treatment Problem," *IEEE Transactions on Systems Science and Cybernetics* (New York: Institute of Electrical and Electronics Engineers, Inc., Sept. 1968), pp. 360-361.

[3]Donald C. Harrison and Norman Shumway, "Evaluation and Surgery for Impending Myocardial Infarction". Stanford University. *The Myocardium*.

[4]BMD05M, *BMD Biomedical Computer Programs* (Los Angeles: University of California Press, 1974), pp. 221-231.

[5]A. A. Afifi and S. P. Azen, *Statistical Analysis: A Computer Oriented Approach* (New York: Academic Press, 1972), pp. 235-252.

BIBLIOGRAPHY

Afifi, A. A. and Azen, S. P. Statistical Analysis: A Computer Oriented Approach. New York: Academic Press, 1972.

Dixon, W. J., Editor. BMD: Biomedical Computer Programs. Los Angeles: University of California Press, 1974.

Ginsberg, A. and Offensend, F. "An Application of Decision Theory to a Medical Diagnosis - Treatment Problem," IEEE Transactions on Systems Science and Cybernetics. New York: Institute of Electric and Electronic Engineers, Inc. September 1968.

Harrison, Donald and Shumway, Norman. "Evaluation and Surgery for Impending Myocardial Infarction." The Myocardium.

Harrison, T. et. al. Principles of Internal Medicine. New York: McGraw Hill Book Company, Inc. 1962.

Levy, H. Annals of Internal Medicine 44: 1123, 1956.

Schroeder, J. et. al. "Clinical Course of Chest Pain Admissions" (unpublished report of Cardiology Division of Stanford University Medical Center, October 1974.)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

6 A Decision Analytic Approach to Coronary Care Patient Classification.

5. TYPE OF REPORT & PERIOD COVERED

9 Final rept.,

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

10 [redacted] Terry A./Bresnick

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Dept of Engineering ✓
West Point, New York

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Dept of Engineering
West Point, New York

12. REPORT DATE

11 Oct [redacted]76

13. NUMBER OF PAGES

~~18~~ 12 23 p.

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Presented at the Joint National Meeting of the Operations Research Society of America/The Institute of Management Science" in Miami, Fla., Nov. 1976

18. SUPPLEMENTARY NOTES

Presented at the Joint National Meeting of the Operations Research Society of America/The Institute of Management Science, Miami Florida, Nov. 1976.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Coronary Care
Decision Analysis
Discriminant Analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this project is to categorize cardiac patients using an expected utility criteria. Bayesian updating using symptomatic information has been a successful tool for disease diagnosis. However, in classification of cardiac patients, discriminant analysis using the BMD05M computer package achieves comparable results with far less effort. Discriminant scores can be evaluated and normalized as probabilistic estimates of the patient's cardiac condition. By incorporating these probabilities with subjective value estimates for possible mis-classifications, the expected utility of assigning patients (con't)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

409 732 del

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20. Abstracts (con't)

to each cardiac category can be calculated. Patients are then classified into cardiac categories based upon highest expected utility.